Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 1084–1093
Expert Systems with Applications
www.elsevier.com/locate/eswa

EEG signal classification using wavelet feature extraction and a mixture of expert model

Abdulhamit Subasi *

Department of Electrical and Electronics Engineering, Kahramanmaras Sutcu Imam University, Avsar Yerleskesi, 46050-9 Kahramanmaraş, Turkey

Abstract

Mixture of experts (ME) is modular neural network architecture for supervised learning. A double-loop Expectation-Maximization (EM) algorithm has been introduced to the ME network structure for detection of epileptic seizure. The detection of epileptiform discharges in the EEG is an important component in the diagnosis of epilepsy. EEG signals were decomposed into the frequency sub-bands using discrete wavelet transform (DWT). Then these sub-band frequencies were used as an input to a ME network with two discrete outputs: normal and epileptic. In order to improve accuracy, the outputs of expert networks were combined according to a set of local weights called the "gating function". The invariant transformations of the ME probability density functions include the permutations of the expert labels and the translations of the parameters in the gating functions. The performance of the proposed model was evaluated in terms of classification accuracies and the results confirmed that the proposed ME network structure has some potential in detecting epileptic seizures. The ME network structure achieved accuracy rates which were higher than that of the stand-alone neural network model.

Keywords: Electroencephalogram (EEG); Epileptic seizure; Discrete wavelet transform (DWT); Mixture of experts; Expectation-Maximization (EM) algorithm

1. Introduction

Temporary electrical disturbance of the brain causes epileptic seizures. Sometimes seizures may go unnoticed, depending on their presentation, and sometimes may be confused with other events, such as a stroke, which can also cause falls or migraines. Approximately one in every 100 persons will experience a seizure at some time in their life (Adeli, Zhou, & Dadmehr, 2003). Unfortunately, the occurrence of an epileptic seizure seems unpredictable and its course of action is very little understood. Research is needed for better understanding of the mechanisms causing epileptic disorders. Careful analysis of the electroencephalograph (EEG) records can provide valuable insight into this widespread brain disorder. The detection of epileptiform discharges occurring in the EEG between seizures is an important component in the diagnosis of epilepsy. (Adeli et al., 2003; Subasi, 2005a).

Spectral analysis of the EEG signals produces information about the brain activities. However, artificial neural networks (ANNs) may offer a potentially superior method of EEG signal analysis to the spectral analysis methods. In contrast to the conventional spectral analysis methods, ANNs not only model the signal, but also make a decision as to the class of signal (Subasi, 2005a; Subasi & Ercelebi, 2005). Neural networks have been successfully used in a various medical applications (Baxt, 1990; Miller, Blott, & Hames, 1992). Recent advances in the field of neural networks have made them attractive for analyzing signals. The application of neural networks has opened a new area for solving problems not resolvable by other signal processing techniques (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya,

* Tel.: +90 344 219 1253; fax: +90 344 219 1052.
E-mail address: asubasi@ksu.edu.tr

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.02.005

2000; Guler & Ubeyli, 2005). In recent times there have been widespread interests in the use of multiple models for pattern classification and regression in statistics and neural network communities. The crucial idea underlying these methods is the application of a so-called divide-and-conquer principle that is frequently used to deal with a complex problem by dividing it into simpler problems whose solutions can be combined to yield a final solution. Utilizing this principle, Jacobs, Jordan, Nowlan, and Hinton (1991) proposed a modular neural network architecture called mixture of experts (ME). The ME network contains a population of simple linear classifiers (the "experts") whose outputs are mixed by a "gating" network. During learning, the experts compete to classify each input training pattern, and the gating network directs more error information (feedback) to the expert that performs best. Eventually, the gating network learns to partition the input space such that expert 1 "specializes" in one area of the space, expert 2 specializes in another area of the space, and so on. As pointed out by Jordan and Jacobs (1994), the gating network performs a typical multi-class classification task. Moreover, Expectation–Maximization (EM) algorithm have been applied to the ME architecture so that the learning process is decoupled in a manner that fits well with the modular structure (Dempster, Laird, & Rubin, 1977; Jordan & Jacobs, 1994). The favorable properties of the EM algorithm have been shown by theoretical analyses (Chen, Xu, & Chi, 1999; Jordan & Xu, 1995; Xu & Jordan, 1996).

The EM algorithm can be extended to provide an effective training mechanism for the MEs based on a Gaussian probability assumption. Although originally the model structure is predetermined and the training algorithm is based on the Gaussian probability assumption for each expert model output, the ME framework is a powerful concept that can be extended to a wide variety of applications including medical diagnostic decision support system applications due to numerous inherent advantages (Chen et al., 1999; Guler & Ubeyli, 2005; Hong & Harris, 2002; Jordan & Jacobs, 1994; Mangiameli & West, 1999).

Until now, there is no study in the literature related to the estimation of ME accuracy in analysis of EEG signals. In this study, a new approach based on ME was presented for epileptic seizure detection. The ME network was used to detect epileptic seizure when statistical features of discrete wavelet transform (DWT) sub-band frequencies were used as inputs. In the configuration of ME for the detection of epileptic seizure, we used two local experts and a gating network, which were in the form of multi-layer perceptron neural networks (MLPNNs), since there were two possible outcomes of the detection of epileptic seizure (epileptic or not). We were able to achieve considerable enhancement in accuracy by applying ME with EM algorithm compared to the stand-alone neural networks. Finally, some conclusions were drawn regarding the impacts of features on epileptic seizure detection.

2. Materials and method

2.1. Data selection and recording

We used the publicly available data described in Andrzejak et al. (2001). In this section, we restrict ourselves to only a short description and refer to Andrzejak et al. (2001) for further details. The complete data set consists of five sets (denoted A–E) each containing 100 single-channel EEG segments. These segments were selected and cut out from continuous multi-channel EEG recordings after visual inspection for artifacts, e.g., due to muscle activity or eye movements. Sets A and B consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme (Fig. 1). Volunteers were relaxed in an awake state with eyes open (A) and eyes closed (B), respectively. Sets C, D, and E originated from EEG archive of presurgical diagnosis. EEGs from five patients were selected, all of whom had achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the epileptogenic zone. Segments in set D were recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity. Here segments were selected from all recording sites exhibiting ictal activity. All EEG signals were recorded with the same 128-channel amplifier system, using an average common reference. The data were digitized at 173.61 samples per second using 12 bit resolution. Band-pass filter settings were 0.53–40 Hz (12 dB/oct). In this study, we used two dataset (A and E) of the complete dataset. Typical EEGs are depicted in Fig. 2.

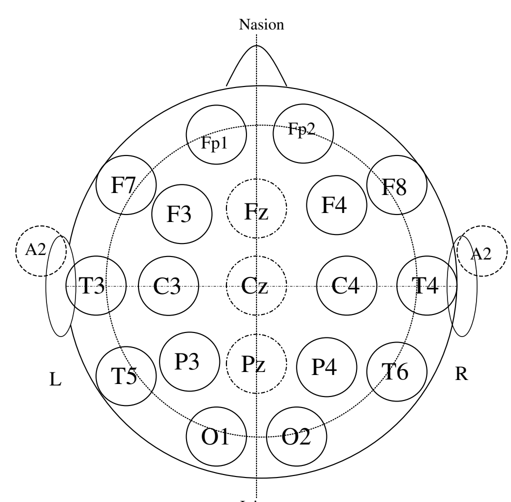

Fig. 1. The 10–20 international system of electrode placement.

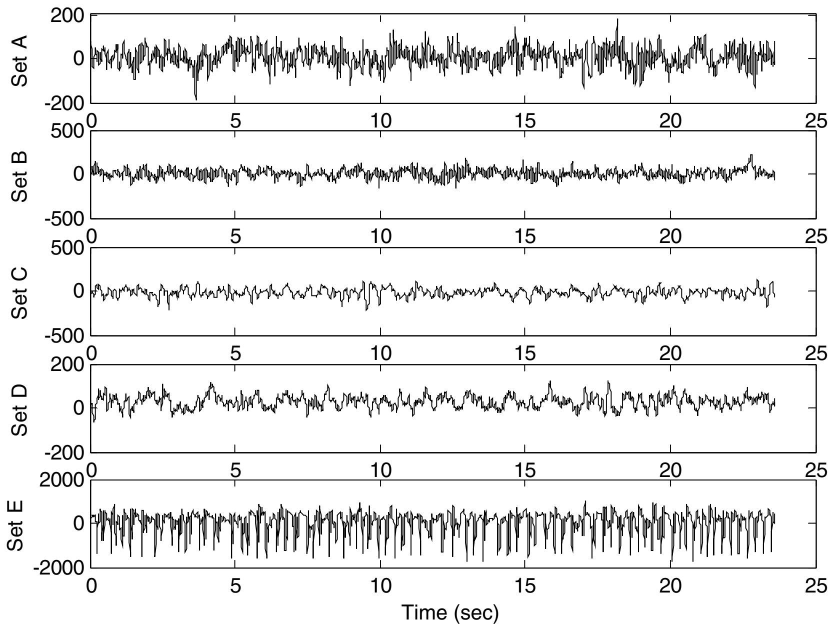

Fig. 2. Examples of five different sets of EEG signals taken from different subjects.

2.2. Analysis using discrete wavelet transform

Wavelet transform is a spectral estimation technique in which any general function can be expressed as an infinite series of wavelets. The basic idea underlying wavelet analysis consists of expressing a signal as a linear combination of a particular set of functions (wavelet transform, WT), obtained by shifting and dilating one single function called a mother wavelet. The decomposition of the signal leads to a set of coefficients called wavelet coefficients. Therefore the signal can be reconstructed as a linear combination of the wavelet functions weighted by the wavelet coefficients. In order to obtain an exact reconstruction of the signal, adequate number of coefficients must be computed. The key feature of wavelets is the time-frequency localization. It means that most of the energy of the wavelet is restricted to a finite time interval. Frequency localization means that the Fourier transform is band limited. When compared to STFT, the advantage of time-frequency localization is that wavelet analysis varies the time-frequency aspect ratio, producing good frequency localization at low frequencies (long time windows), and good time localization at high frequencies (short time windows). This produces a segmentation, or tiling of the time-frequency plane that is appropriate for most physical signals, especially those of a transient nature. The wavelet technique applied to the EEG signal will reveal features related to the transient nature of the signal which are not obvious by the Fourier transform. In general, it must be said that no time-frequency regions but rather time-scale regions are defined (Subasi, 2005a, 2005b; Subasi & Ercelebi, 2005).

All wavelet transforms can be specified in terms of a low-pass filter g, which satisfies the standard quadrature mirror filter condition

$$G(z)G(z^{-1}) + G(-z)G(-z^{-1}) = 1, \tag{1}$$

where $G(z)$ denotes the z-transform of the filter g. Its complementary high-pass filter can be defined as

$$H(z) = zG(-z^{-1}). \tag{2}$$

A sequence of filters with increasing length (indexed by i) can be obtained

$$\begin{aligned} G_{i+1}(z) &= G(z^{2^i})G_i(z), \\ H_{i+1}(z) &= H(z^{2^i})G_i(z), \quad i = 0, \ldots, I-1, \end{aligned} \tag{3}$$

with the initial condition $G_0(z) = 1$. It is expressed as a two-scale relation in time domain

$$g_{i+1}(k) = [g]_{\uparrow 2^i} g_i(k), h_{i+1}(k) = [h]_{\uparrow 2^i} g_i(k), \tag{4}$$

where the subscript $[\cdot]_{\uparrow m}$ indicates the up-sampling by a factor of m and k is the equally sampled discrete time.

The normalized wavelet and scale basis functions $\varphi_{i,l}(k)$, $\psi_{i,l}(k)$ can be defined as

$$\begin{aligned} \varphi_{i,l}(k) &= 2^{i/2} g_i(k - 2^i l), \\ \psi_{i,l}(k) &= 2^{i/2} h_i(k - 2^i l), \end{aligned} \tag{5}$$

where the factor $2^{i/2}$ is an inner product normalization, i and l are the scale parameter and the translation parameter, respectively. The DWT decomposition can be described as

$$\begin{aligned} a_{(i)}(l) &= x(k)^* \varphi_{i,l}(k), \\ d_{(i)}(l) &= x(k)^* \psi_{i,l}(k), \end{aligned} \tag{6}$$

where $a_{(i)}(l)$ and $d_i(l)$ are the approximation coefficients and the detail coefficients at resolution i, respectively (Akay, 1997; Guler & Ubeyli, 2005).

The discrete wavelet transform (DWT) is an adaptable signal processing tool that finds many engineering and scientific applications. One area in which the DWT has been particularly successful is the epileptic seizure detection because it captures transient features and localizes them in both time and frequency content accurately. DWT analyzes the signal at different frequency bands, with different resolutions by decomposing the signal into a coarse approximation and detail information. DWT employs two sets of functions called scaling functions and wavelet functions, which are related to low-pass and high-pass filters, respectively. The decomposition of the signal into the different frequency bands is merely obtained by consecutive high-pass and low-pass filtering of the time domain signal. The procedure of multi-resolution decomposition of a signal $x[n]$ is schematically shown in Fig. 3. Each stage of this scheme consists of two digital filters and two downsamplers by 2. The first filter, $h[\cdot]$ is the discrete mother wavelet, high-pass in nature, and the second, $g[\cdot]$ is its mirror version, low-pass in nature. The down-sampled outputs of first high-pass and low-pass filters provide the detail, D_1 and the approximation, A_1, respectively. The first approximation, A_1 is further decomposed and this process is continued as shown in Fig. 3 (Subasi, 2005b).

Selection of suitable wavelet and the number of decomposition levels is very important in analysis of signals using the DWT. The number of decomposition levels is chosen based on the dominant frequency components of the signal. The levels are chosen such that those parts of the signal that correlate well with the frequencies necessary for classification of the signal are retained in the wavelet coefficients. In the present study, since the EEG signals do not have any useful frequency components above 30 Hz, the number of decomposition levels was chosen to be 5. Thus, the EEG signals were decomposed into details D1–D5 and one final approximation, A5. Usually, tests are performed with different types of wavelets and the one which gives maximum efficiency is selected for the particular application. The smoothing feature of the Daubechies wavelet of order 4 (db4) made it more appropriate to detect changes of EEG signals. Hence, the wavelet coefficients were computed using the db4 in the present study. In order to investigate the effect of other wavelets on classifications accuracy, tests were carried out using other wavelets also. Apart from db4, Symmlet of order 10 (sym10), Coiflet of order 4 (coif4), and Daubechies of order 2 (db2) were also tried. It was noticed that the Daubechies wavelet gives better accuracy than the others, and db4 is slightly better than db2.

The proposed method was applied on both data set of EEG data (Sets A and E). Fig. 4 shows approximation (A5) and details (D1–D5) of an epileptic EEG signal. Fig. 5 shows approximation (A5) and details (D1–D5) of a normal EEG signal. These approximation and detail records are reconstructed from the Daubechies 4 (DB4) wavelet filter. Wavelet transform acts like a mathematical microscope, zooming into small scales to reveal compactly spaced events in time and zooming out into large scales to exhibit the global waveform patterns (Adeli et al., 2003).

2.3. Feature extraction

The extracted wavelet coefficients provide a compact representation that shows the energy distribution of the EEG signal in time and frequency. Table 1 presents frequencies corresponding to different levels of decomposition for Daubechies order 4 wavelet with a sampling frequency of 173.6 Hz. In order to further decrease the dimensionality of the extracted feature vectors, statistics over the set of the wavelet coefficients was used (Kandaswamy, Kumar,

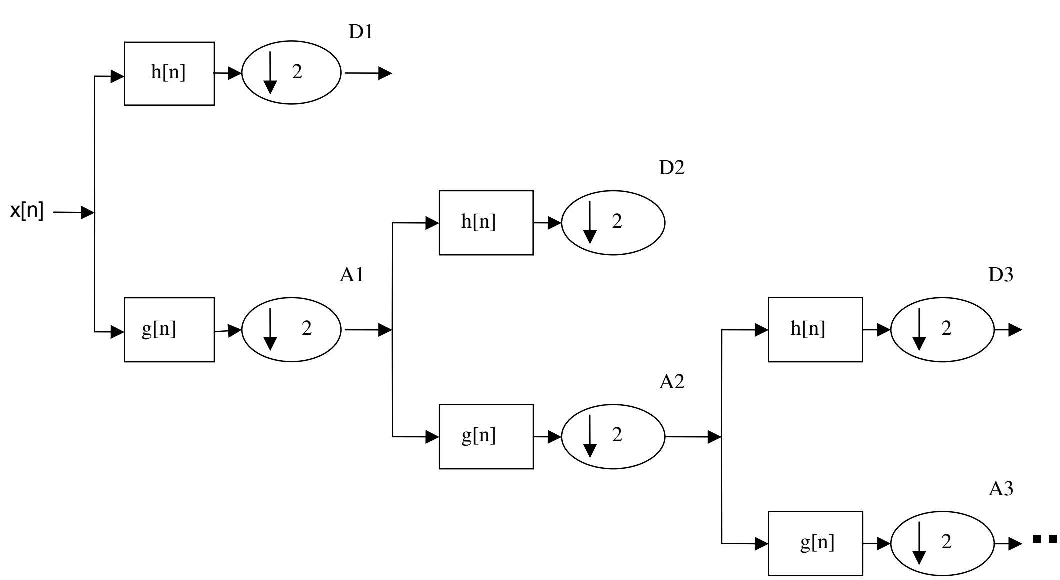

Fig. 3. Sub-band decomposition of DWT implementation; $h[n]$ is the high-pass filter, $g[n]$ the low-pass filter.

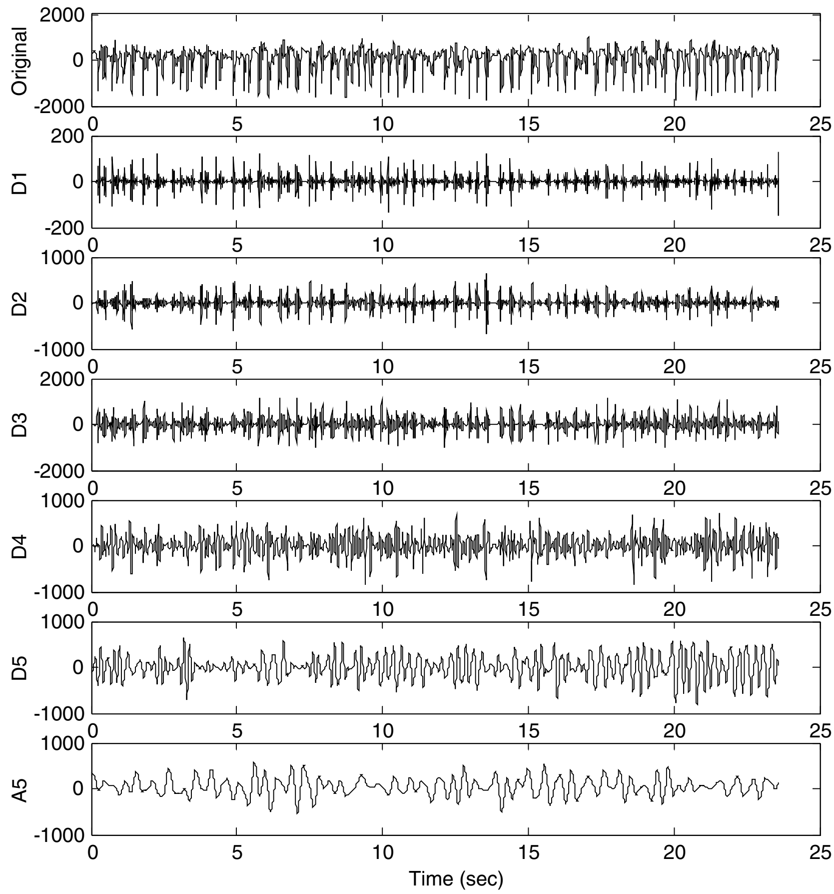

Fig. 4. Approximate and detailed coefficients of EEG signal taken from unhealthy subject (epileptic patient).

Ramanathan, Jayaraman, & Malmurugan, 2004). The following statistical features were used to represent the time-frequency distribution of the EEG signals:

(1) Mean of the absolute values of the coefficients in each sub-band.
(2) Average power of the wavelet coefficients in each sub-band.
(3) Standard deviation of the coefficients in each sub-band.
(4) Ratio of the absolute mean values of adjacent sub-bands.

Features 1 and 2 represent the frequency distribution of the signal and the features 3 and 4 the amount of changes in frequency distribution. These feature vectors, calculated for the frequency bands A5 and D3–D5, were used for classification of the EEG signals.

2.4. Artificial neural network models

Artificial neural networks (ANNs) are computing systems made up of large number of firmly interconnected adaptive processing elements (neurons) that are able to perform massively parallel computations for data processing and knowledge representation. Learning in ANNs is accomplished through special training algorithms developed based on learning rules presumed to mimic the learning mechanisms of biological systems. ANNs can be trained to recognize patterns and the nonlinear models developed during training allow neural networks to generalize their conclusions and to make application to patterns not previously encountered (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Guler & Ubeyli, 2005; Haykin, 1994).

2.4.1. Multi-layer perceptron neural networks (MLPNN)

The MLPNNs, which have features such as the ability to learn and generalize, smaller training set requirements, fast operation, ease of implementation and therefore most commonly used neural network architectures, have been adapted for describing the alertness level of arbitrary subject. Currently, the most widely used ANN type is a MLPNN which has been playing a central role in applications of neural networks. The MLPNN is a nonparametric technique for performing a wide variety of detection and estimation tasks (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Haykin, 1994). In the MLPNN, each neuron j in the hidden layer sums its input signals x_i after multiplying them by the strengths of the respective

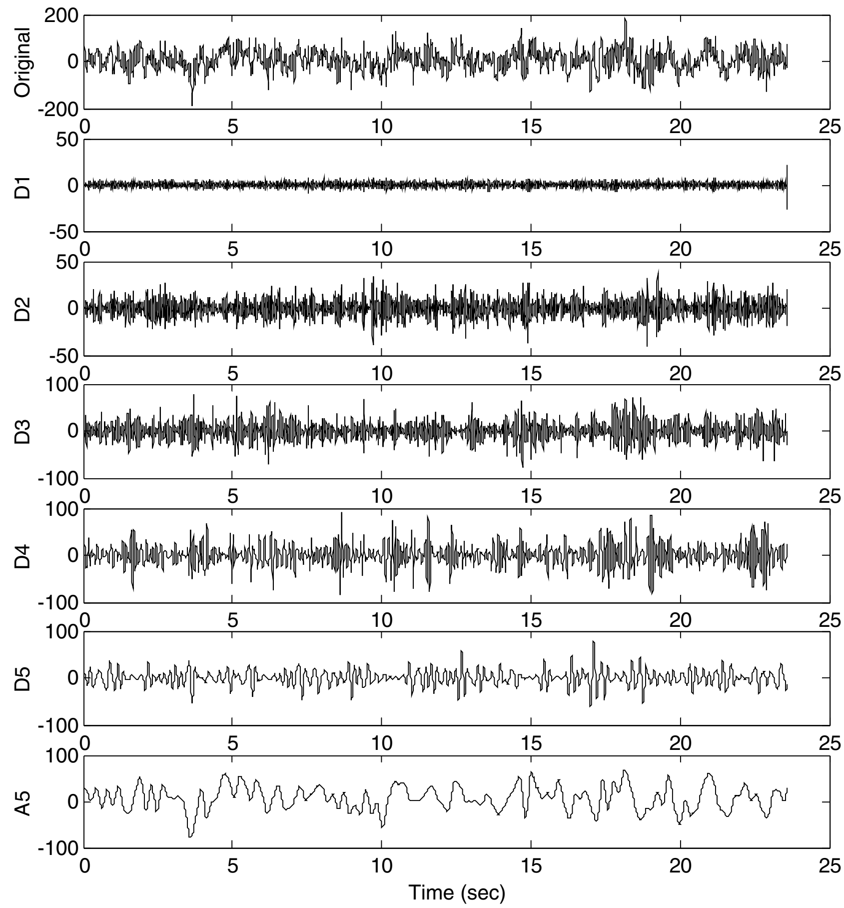

Fig. 5. Approximate and detailed coefficients of EEG signal taken from a healthy subject.

Table 1
Frequencies corresponding to different levels of decomposition for Daubechies 4 filter wavelet with a sampling frequency of 173.6 Hz

Decomposed signal	Frequency range (Hz)
D1	43.4–86.8
D2	21.7–43.4
D3	10.8–21.7
D4	5.4–10.8
D5	2.7–5.4
A5	0–2.7

connection weights w_{ji} and computes its output y_j as a function of the sum

$$y_j = f\left(\sum w_{ji}x_i\right), \tag{7}$$

where f is the activation function that is essential to transform the weighted sum of all signals impinging onto a neuron. The activation function (f) can be a simple threshold function, or a sigmoid, hyperbolic tangent, or radial basis function. The sum of squared differences between the desired and actual values of the output neurons E is defined as

$$E = \frac{1}{2}\sum_j \left(y_{dj} - y_j\right)^2, \tag{8}$$

where y_{dj} is the desired value of output neuron j and y_j is the actual output of that neuron. Each weight w_{ji} is adjusted to reduce E as rapidly as possible. How w_{ji} is adjusted depends on the training algorithm adopted (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Haykin, 1994).

Training algorithms are a primary part of ANN model development. A suitable topology may still fail to give a better model, unless trained by a suitable training algorithm. A good training algorithm will shorten the training time, while achieving a better accuracy. Therefore, training process is an important characteristic of the ANNs, whereby representative examples of the knowledge are iteratively presented to the network, so that it can integrate this knowledge within its structure. There are a number of training algorithms used to train a MLPNN and a frequently used one is called the backpropagation training algorithm (Basheer & Hajmeer, 2000; Chaudhuri & Bhattacharya, 2000; Guler & Ubeyli, 2004; Haykin, 1994). The backpropagation algorithm, which is based on searching an error surface using gradient descent for points with minimum error, is relatively easy to implement. However, backpropagation has some problems for many applications. The algorithm is not guaranteed to find the global minimum of the error function since gradient descent may get stuck in local minima, where it may remain indefinitely. In addition to this, long training sessions are often required in order to find an acceptable weight solution because of the well-known difficulties inherent in gradient descent optimization. Therefore, a lot of variations to improve the convergence of the backpropagation were proposed.

2.4.2. Mixtures of experts and EM algorithm

In this section we present ME, a method for epileptic seizure detection based on a mixture of expert models. ME focuses on the problem of learning a mapping in which the form of the mapping is different for different regions of the input space. Although a single homogeneous adaptive model could be applied to this problem, we might expect that the task would be better performed if we assign different "expert" models to tackle each of the different regions, and then use an extra "gating" model, which also checks the input vector, to make a decision which one of the experts should be used to find out the output. If the problem has an apparent decomposition of this form, then it may be feasible to design the ME system by hand. However, a more powerful and more general approach would be to realize a suitable decomposition as part of the learning process. Our realization of the gating model in ME is based on the clustering algorithm. The idea of ME is "to divide for conquer". A complex problem is subdivided into simpler sub problems that are treated independently (Milidiu, Machado, & Renteria, 1999).

As seen in Fig. 6, the ME architecture is consists of a gating network and several expert networks. The gating network receives the vector x as input and produces scalar outputs that are partition of unity at each point in the input space. Each expert network produces an output vector for an input vector. The gating network provides linear combination coefficients as veridical probabilities for expert networks and, therefore, the final output of the ME architecture is a convex weighted sum of all the output vectors produced by expert networks. Suppose that there are N expert networks in the ME architecture. All the expert

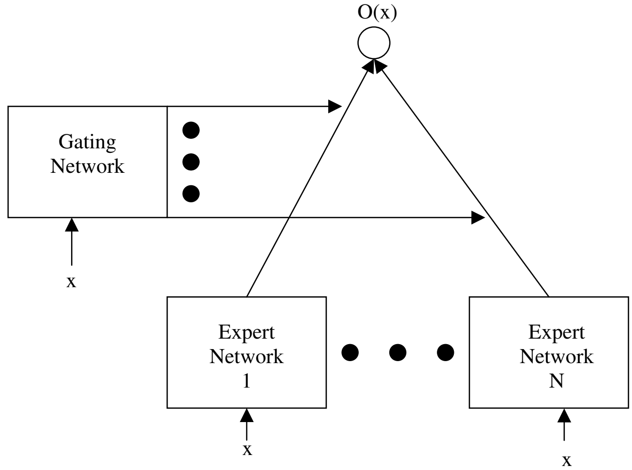

Fig. 6. The architecture of mixture of experts.

networks are linear with a single output nonlinearity that is also referred to as "generalized linear" (McCullagh & Nelder, 1983). The ith expert network produces its output $o_i(x)$ as a generalized linear function of the input x

$$o_i(x) = f(W_i x), \tag{9}$$

where W_i is a weight matrix and $f(\cdot)$ is a fixed continuous nonlinearity. The gating network is also generalized linear function, and its ith output, $g(x, v_i)$, is the multinomial logit or softmax function of intermediate variables ζ_i (Bridle, 1989; Chen et al., 1999; McCullagh & Nelder, 1983):

$$g(x, v_i) = \frac{e^{\xi_i}}{\sum_{k=1}^{N} e^{\xi_k}} \tag{10}$$

where $\xi_i = v_i^T x$ and v_i is a weight vector. The overall output $o(x)$ of the ME architecture is

$$o(x) = \sum_{k=1}^{N} g(x, v_k) o_k(x). \tag{11}$$

The ME architecture can be given a probabilistic interpretation. For an input–output pair (x, y), the values of $g(v_i, x)$ are interpreted as the multinomial probabilities associated with the decision that terminates in a regressive process that maps x to y. Once the decision has been made, resulting in a choice of regressive process i, the output y is then chosen from a probability density $P(y|x, W_i)$, where W_i denotes the set of parameters or weight matrix of the ith expert network in the model. Therefore, the total probability of generating y from x is the mixture of the probabilities of generating y from each component densities, where the mixing proportions are multinomial probabilities

$$P(y|x, \varphi) = \sum_{k=1}^{N} g(x, v_k) P(y|x, W_k), \tag{12}$$

where φ is the set of all the parameters including both expert and gating network parameters. Moreover, the probabilistic component of the model is generally assumed to be a Gaussian distribution in the case of regression, a Bernoulli distribution in the case of binary classification, and a multinomial distribution in the case of multi-class classification (Chen et al., 1999).

Based on the probabilistic model in Eq. (12), learning in the ME architecture is treated as a maximum likelihood problem. Jordan and Jacobs (1994) have proposed an EM algorithm for adjusting the parameters of the architecture. Suppose that the training set is given as: $X = \{(x_1, y_1)\}_{t=1}^{T}$. The EM algorithm consists of two steps. For the sth epoch, the posterior probabilities $h_i(t)$ $(i = 1, \ldots, N)$, which can be interpreted as the probabilities $P(i|x_t, y_t)$, are computed in the E-step as

$$h_i^{(t)} = \frac{g(x_t, v_i^{(s)}) P(y_t|x_t, W_i^{(s)})}{\sum_{k=1}^{N} g(x_t, v_k^{(s)}) P(y_t|x_t, W_k^{(s)})} \tag{13}$$

The M-step solves the following maximization problems:

$$W_i^{(s+1)} = \arg\max_{W_i} \sum_{t=1}^{T} h_i^{(t)} \log P(y_t|x_t, W_i), \tag{14}$$

and

$$V^{(s+1)} = \arg\max_{V} \sum_{t=1}^{T} \sum_{k=1}^{N} h_k^{(t)} \log g(x_t, v_k), \tag{15}$$

where V is the set of all the parameters in the gating network. Therefore, the EM algorithm (Chen et al., 1999) is summarized as

EM algorithm

1. For each data pair (x_t, y_t), compute the posterior probabilities $h_i(t)$ using the current values of the parameters.
2. For each expert network i, solve a maximization problem in Eq. (14) with observations $\{(x_t, y_t)\}_{t=1}^{T}$, and observation weights $\{h_i^{(t)}\}_{t=1}^{T}$.
3. For the gating network, solve the maximization problem in Eq. (15) with observations $\{(x_t, h_k^{(t)})\}_{t=1}^{T}$.
4. Iterate by using the updated parameter values.

2.5. Statistical measures

The statistical measures used to evaluate the performance were:

True positives (TP): the number of seizures identified by the automated system and by the EEG experts.
False positives (FP): the events identified as seizures by the automated system but not by the EEG experts.
False negatives (FN): the events identified as seizures by the experts but missed by the automated system.

The sensitivity value (true positive, same positive result as the diagnosis of expert neurologists) was calculated by dividing the total of diagnosis numbers to total diagnosis numbers that are stated by the expert neurologists. Sensitivity, also called the true positive ratio, is calculated by the formula

$$\text{Sensitivity} = \text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\%. \tag{16}$$

On the other hand, specificity value (true negative, same diagnosis as the expert neurologists) is calculated by dividing the total of diagnosis numbers to total diagnosis numbers that are stated by the expert neurologists. Specificity, also called the true negative ratio, is calculated by the formula

$$\text{Specificity} = \text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%. \tag{17}$$

3. Results and discussion

The basic requirement to find an accurate model is the collection of well distributed, sufficient, and accurately measured input data. The key component of designing the classifier based on pattern classification is choice of the ME inputs, because even the best classifier will perform inadequately if the inputs are not selected well. Input selection has two meanings: (1) which components of a pattern, or (2) which set of inputs best represent a given pattern. In this study, EEG recordings were divided into sub-band frequencies such as δ (approximate wavelet coefficient A5), θ (detailed wavelet coefficient D5), α (detailed wavelet coefficient D4) and β (detailed wavelet coefficient D3) by using DWT (Figs. 4 and 5). Then a set of statistical features was extracted from the wavelet sub-band frequencies. After normalization, the EEG signals were decomposed using wavelet transform and the statistical features were extracted from the sub-bands. A classification system based on ME was implemented using the statistical features as inputs.

The ME architecture used for the detection of epileptic seizure is shown in Fig. 6. Since we investigated two-group classification completely, the ME was configured with two local experts and a gating network which were in the form of MLPNNs. The computed discrete wavelet coefficients were used as the inputs of the MLPNNs employed in the architecture of ME. In order to extract features, the wavelet coefficients corresponding to the D1–D5 frequency bands of the EEG signals were computed. In order to reduce the dimensionality of the extracted feature vectors, statistics explained in Section 2.3 over the set of the wavelet coefficients was used.

The objective of the modelling phase in this application was to develop classifiers that are able to identify any input combination as belonging to either one of the two classes: normal or epileptic. For developing neural network classifiers, 1000 examples were randomly taken from the 1600 examples and used for training the neural networks, and the remaining 600 examples were kept aside and used for testing the developed models. The class distribution of the samples in the training and validation data set is summarized in Table 2.

In classification, we want to assign the input patterns to one of several classes, usually represented by outputs restricted to lie in the range from 0 to 1, so that they represent the probability of class membership. While the classification is carried out, a specific pattern is assigned to a specific class according to the characteristic features

Table 2
Class distribution of the samples in the training and test data sets

Class	Training set	Test set	Total
Epileptic	500	300	800
Normal	500	300	800
Total	1000	600	1600

selected for it. In this application, there were two classes: epileptic or healthy. Classification results of the ME were displayed by a confusion matrix. The confusion matrix showing the classification results of the ME is given below.

Result	Epileptic	Normal
Confusion matrix		
Epileptic	285	15
Normal	18	282

Table 3
The values of statistical parameters of the ME and MLPNN models for EEG signal classification

Classifier type	Correctly classified (%)	Specificity (%)	Sensitivity (%)
ME	94.5	94	95
MLPNN	93.2	92.6	93.6

According to the confusion matrix, 18 healthy subjects were classified incorrectly by the ME as an epileptic patient, 15 epileptic patients were classified as a normal subject. The test performance of the ME was determined by the computation of the statistical parameters given in 2.5.

Firstly we used statistical features extracted from the sub-bands of EEG signals for MLPNN and ME classification. The procedure was repeated on EEG recordings of all subjects (healthy and epileptic patients). Table 3 shows a summary of the performance measures by using statistical features extracted from the sub-bands of EEG signals using DWT. It is obvious from Table 3 that the ME classifier is ranked first in terms of its correct classification percentage of the EEG signals (epileptic/normal data 94.5%), while the MLPNN-based classifier came second (93.2%). The ME classifier identified accurately all the epileptic and normal cases with specificity 94% and sensitivity 95%, followed by the MLPNN-based classifier with specificity 92.6% and sensitivity 93.6%.

The testing performance of the ME network diagnostic system is found to be satisfactory and we think that this system can be used in clinical studies in the future after it is developed. This application brings objectivity to the evaluation of EEG signals and its automated nature makes it easy to be used in clinical practice. Besides the feasibility of a real-time implementation of the expert diagnosis system, diagnosis may be made more accurately by increasing the variety and the number of parameters. A "black box" device that may be developed as a result of this study may provide feedback to the neurologists for classification of the EEG signals quickly and accurately by examining the EEG signals with real-time implementation.

4. Concluding remarks

It is a difficult task to diagnose epilepsy and requires observation of the patient, an EEG, and gathering of additional clinical information. An artificial neural network that classifies subjects as having or not having an epileptic seizure provides a valuable diagnostic decision support tool for physicians treating potential epilepsy, since differing etiologies of seizures result in different treatments.

In this paper, two approaches to develop classifiers for identifying epileptic seizure were discussed. One approach is based on the traditional neural network technology, mainly using MLPNN and the other is the Mixture of Expert (ME). Using statistical features extracted from the DWT sub-bands of EEG signals, ME and MLPNN were constructed and cross-compared in terms of their accuracy relative to the observed epileptic/normal patterns. The comparisons were based on two scalar performance measures derived from the confusion matrices; namely specificity and sensitivity. Out of the 600 epileptic/normal cases, the MLPNN-based classifier misclassified 41 cases, while the ME-based classifier misclassified 33 cases.

In this work we have presented the use of ME network structures to improve accuracy of epileptic seizure detection in EEG since the overall structure predictive performance is generally superior to any of the individual experts. In order to detect the epileptic seizure in EEG, two local experts and a gating network, which were in the form of MLPNNs, were used in the configuration of ME architecture. EM algorithm was used for training the ME networks so that the learning process is decoupled in a manner that fits well with the modular structure. The ME used for the epileptic seizure detection was trained, cross validated and tested with the extracted features using DWT of the EEG signals obtained from healthy and unhealthy subjects. The classification results and the values of statistical parameters were used for evaluating performances of the classifiers. The accuracy rates achieved by the ME network structures presented for the epileptic seizure detection were found to be higher than that of the stand-alone neural network model. With specificity and sensitivity values both above 94%, the wavelet neural network classification may be used as an important diagnostic decision support mechanism to assist physicians in the treatment of epileptic patients.

References

Adeli, H., Zhou, Z., & Dadmehr, N. (2003). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods, 123*, 69–87.

Akay, M. (1997). Wavelet applications in medicine. *IEEE Spectrum, 34*(5), 50–56.

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E, 64*, 061907.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods, 43*, 3–31.

Baxt, W. G. (1990). Use of an artificial neural network for data analysis in clinical decision making: the diagnosis of acute coronary occlusion. *Neural Computation, 2*, 480–489.

Bridle, J. (1989). Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In F. Fogelman-Soulie & J. Herault (Eds.), *Neurocomputing: algorithm, architectures, and applications* (pp. 227–236). New York: Springer.

Chaudhuri, B. B., & Bhattacharya, U. (2000). Efficient training and improved performance of multilayer perceptron in pattern classification. *Neurocomputing, 34*, 11–27.

Chen, K., Xu, L., & Chi, H. (1999). Improved learning algorithms for mixture of experts in multiclass classification. *Neural Networks, 12*(9), 1229–1252.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B, 39*(1), 1–38.

Guler, N. F., & Ubeyli, E. D. (2004). Wavelet-based neural network analysis of ophthalmic artery Doppler signals. *Computers in Biology and Medicine, 34*(7), 601–613.

Guler, I., & Ubeyli, E. D. (2005). A mixture of experts network structure for modelling Doppler ultrasound blood flow signals. *Computers in Biology and Medicine, 35*, 565–582.

Haykin, S. (1994). *Neural networks: A comprehensive foundation*. New York: Macmillan.

Hong, X., & Harris, C. J. (2002). A mixture of experts network structure construction algorithm for modelling and control. *Applied Intelligence, 16*(1), 59–69.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3*(1), 79–87.

Jordan, M. I., & Jacobs, R. A. (1994). Hierarchical mixture of experts and the EM algorithm. *Neural Computation, 6*(2), 181–214.

Jordan, M. I., & Xu, L. (1995). Convergence results for the EM approach to mixtures of experts. *Neural Networks, 8*(9), 1409–1431.

Kandaswamy, A., Kumar, C. S., Ramanathan, R. P., Jayaraman, S., & Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine, 34*(6), 523–537.

Mangiameli, P., & West, D. (1999). An improved neural classification network for the two-group problem. *Computers and Operations Research, 26*(5), 443–460.

McCullagh, P., & Nelder, J. A. (1983). *Generalized linear models*. London: Chapman and Hall.

Milidiu, R. L., Machado, R. J., & Renteria, R. P. (1999). Time-series forecasting through wavelets transformation and a mixture of expert models. *Neurocomputing, 28*, 145–156.

Miller, A. S., Blott, B. H., & Hames, T. K. (1992). Review of neural network applications in medical imaging and signal processing. *Medical and Biological Engineering and Computing, 30*, 449–464.

Subasi, A. (2005a). Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert Systems with Applications, 28*, 701–711.

Subasi, A. (2005b). Epileptic seizure detection using dynamic wavelet network. *Expert Systems with Applications, 29*, 343–355.

Subasi, A., & Ercelebi, E. (2005). Classification of EEG signals using neural network and logistic regression. *Computer Methods and Programs in Biomedicine, 78*, 87–99.

Xu, L., & Jordan, M. I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Computation, 8*(2), 129–151.